Homework 6

Christophe Hunt

March 12, 2017

Contents

1 Problem Set 1

1.1 (1) When you roll a fair die 3 times, how many possible outcomes are there?

1.2 (2) What is the probability of getting a sum total of 3 when you roll a die two times?

1.3 (3) Assume a room of 25 strangers. What is the probability that two of them have the same birthday? Assume that all birthdays are equally likely and equal to 1/365 each. What happens to this probability when there are 50 people in the room?

2 Problem Set 2

Sometimes you cannot compute the probability of an outcome by measuring the sample space and examining the symmetries of the underlying physical phenomenon, as you could do when you rolled die or picked a card from a shu???ed deck. You have to estimate probabilities by other means. For instance, when you have to compute the probability of various english words, it is not possible to do it by examination of the sample space as it is too large. You have to resort to empirical techniques to get a good enough estimate. One such approach would be to take a large corpus of documents and from those documents, count the number of occurrences of a particular character or word and then base your estimate on that.

Write a program to take a document in English and print out the estimated probabilities for each of the words that occur in that document. Your program should take in a ???le containing a large document and write out the probabilities of each of the words that appear in that document. Please remove all punctuation (quotes, commas, hyphens etc) and convert the words to lower case before you perform your calculations.

Extend your program to calculate the probability of two words occurring adjacent to each other. It should take in a document, and two words (say the and for) and compute the probability of each of the words occurring in the document and the joint probability of both of them occurring together. The order of the two words is not important.

Use the accompanying document for your testing purposes. Compare your probabilities of various words with the Time Magazine corpus: http://corpus.byu.edu/time/